## CLAIMS

What is claimed is:

1. A method for scoring peptide matches, the method comprising:

providing a first peptide and a second peptide;

generating a stochastic model based on one or more match characteristics associated with each of the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide;

calculating a first probability that the first peptide matches the second peptide, based on the stochastic model;

calculating a second probability that the first peptide does not match the second peptide, based on the stochastic model; and

scoring a match between the first peptide and the second peptide based at least in part on a ratio between the first probability and the second probability.

2. The method according to claim 1, where

the first peptide is an experimental peptide; and

the second peptide is a candidate peptide, where the candidate peptide is selected from a group consisting of experimental peptides, theoretical peptides, and a library of peptides.

3. The method according to claim 1, where the one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide comprise at least one of:

mass error;

charge state;

amino acid composition;

missed cleavage;

elution time;

protein/peptide modification;

mass spectrum peak intensity

mass spectrum signal to noise ratio;

mass spectrum signal quality indicator;

statistics derived from a database; and

any observable or derivable characteristics.

4. The method according to claim 1 further comprising determining the probability distributions for the one or more match characteristics.

5. The method according to claim 4 further comprising determining an empirical probability distribution for the one or more match characteristics based on matches between experimental data for known peptides and peptides in a peptide database.

6. The method according to claim 1 further comprising adjusting the stochastic model and a plurality of parameters associated with the stochastic model based on a learning data set, where the learning data set comprises a plurality of peptides that have been identified or a set of known protein standards.

7. The method according to claim 1 further comprising generating an output, where the output comprises at least one of:

a match score associated with the second peptide, where the match score comprises at least one of

a likelihood ratio, where the likelihood ratio is the ratio between the first probability and the second probability;

a log-likelihood, where the log-likelihood is the logarithm of the likelihood ratio;

the likelihood ratio divided by the length of the first peptide measured in amino acids;

the log-likelihood divided by the length of the first peptide measured in amino acids; and

the log-likelihood divided by the logarithm of the length of the first peptide measured in amino acids;

a Z-score associated with the match score;

a p-value associated with the match score;

biological information associated with the first peptide; and

biological information associated with the second peptide.

8. The method according to claim 1, where a theoretical fragmentation spectrum is provided for the second peptide.

9. The method according to claim 8, where the theoretical fragmentation spectrum includes masses corresponding to fragment isotopes.

10. The method according to claim 1 further comprising filtering the second peptide based on at least one of:

the taxonomy of the protein that the second peptide belongs to;

the isoelectric point of the protein that the second peptide belongs to;

the molecular weight of the protein that the second peptide belongs to;

a non-symmetric mass window; and

a set of possible masses made of the union of a plurality of mass intervals.

11. The method according to claim 1 further comprising

providing a physical sample of the first peptide and biological information associated with the first peptide; and

providing a physical sample of the second peptide and biological information associated with the second peptide.

12. A method for scoring a match of two peptides, the method comprising:

providing information associated with an experimental peptide, where the information comprises at least mass spectrum information associated with the experimental peptide and at least one fragment of the experimental peptide;

providing information associated with a candidate peptide;

defining an extended match $E$ based on the information associated with the experimental peptide and the information associated with the candidate peptide;

generating a stochastic model based on the information associated with the experimental peptide and the information associated with the candidate peptide; and

scoring the extended match $E$ based on a likelihood ratio $L = \frac{\mathbf{P}(E|D,s,H_1)}{\mathbf{P}(E|D,s,H_0)}$,

where

$D$ is any extra information that is associated with the experimental peptide and the candidate peptide;

$s$ is a peptide sequence;

$H_1$ is a hypothesis that the peptide sequence $s$ is the correct sequence of the experimental peptide;

$H_0$ is a null-hypothesis that the peptide sequence $s$ is an erroneous sequence of the experimental peptide; and

probabilities $\mathbf{P}(E|D,s,H_1)$ and $\mathbf{P}(E|D,s,H_0)$ are calculated based on the stochastic model.

13. The method according to claim 12, where the extended match $E$ is a random variable that further comprises one or more random variables, the one or more random variables comprising at least one of:

peptide match $P$ that characterizes a match between the experimental peptide mass $m$ and the candidate peptide mass $m_t$;

fragment match $F$ that characterizes a match between fragment masses $f_j$ of the experimental peptide and fragment masses $m_{t,j}$ of the candidate peptide, where $j$ is an index for the fragment masses of the experimental peptide;

charge $z$ that is used to match the $m/z$ ratio of the experimental peptide with the candidate peptide;

elution time $t$ of the experimental peptide;

number of missed cleavages $k$ in the candidate peptide matching the experimental peptide;

protein/peptide modifications $W$ made to the candidate peptide to match the experimental peptide; and

any random variables observable or derivable based on the information associated with the experimental peptide and the candidate peptide.

14. The method according to claim 13 further comprising determining the probability distributions for the one or more random variables.

15. The method according to claim 14 further comprising determining an empirical probability distribution for the one or more random variables based on matches between experimental data for known peptides and peptides in a peptide database.

16. The method according to claim 12 further comprising estimating the probabilities $\mathbf{P}(E|D,s,H_1)$ and $\mathbf{P}(E|D,s,H_0)$ based on the lemma $\mathbf{P}(A,B|C) = \mathbf{P}(A|B,C)\ \mathbf{P}(B|C)$, where $A$, $B$ and $C$ are random variables.

17. The method according to claim 12 further comprising calculating at least one of:

Bayesian score $L' = \dfrac{\mathbf{P}(H_1|D,s,E)}{\mathbf{P}(H_0|D,s,E)} = L\dfrac{\mathbf{P}(H_1|D,s)}{\mathbf{P}(H_0|D,s)}$; and

Bayesian score $L'' = L\dfrac{\mathbf{P}(H_1|D,s,Q)}{\mathbf{P}(H_0|D,s,Q)}$, where $Q$ represents statistics associated with mass spectrum quality of the experimental peptide.

18. The method according to claim 12 further comprising:

comparing the candidate peptide mass with the experimental peptide mass; and

scoring the extended match $E$ based on the likelihood ratio $L$, if the difference between the candidate peptide mass and the experimental peptide mass is in a predetermined range.

19. The method according to claim 12 further comprising adjusting the stochastic model and a plurality of parameters associated with the stochastic model based on a learning data set, where the learning data set comprises a plurality of peptides that have been identified or a set of known protein standards.

20. The method according to claim 12 further comprising generating an output, where the output comprises at least one of:

a match score associated with the candidate peptide, where the match score comprises at least one of

the likelihood;

a log-likelihood, where the log-likelihood is the logarithm of the likelihood ratio;

the likelihood ratio divided by the length of the experimental peptide measured in amino acids;

the log-likelihood divided by the length of the experimental peptide measured in amino acids; and

the log-likelihood divided by the logarithm of the length of the experimental peptide measured in amino acids;

a Z-score associated with the match score;

a p-value associated with the match score;

biological information associated with the experimental peptide; and

biological information associated with the candidate peptide.

21. The method according to claim 12, where a theoretical fragmentation spectrum is provided for the candidate peptide.

22. The method according to claim 21, where the theoretical fragmentation spectrum includes masses corresponding to fragment isotopes.

23. The method according to claim 12 further comprising filtering the candidate peptide based on at least one of:

the taxonomy of the protein that the candidate peptide belongs to;

the isoelectric point of the protein that the candidate peptide belongs to;

the molecular weight of the protein that the candidate peptide belongs to;

a non-symmetric mass window; and

a set of possible masses made of the union of a plurality of mass intervals.

24. The method according to claim 12 further comprising

providing a physical sample of the experimental peptide and biological information associated with the experimental peptide; and

providing a physical sample of the candidate peptide and biological information associated with the candidate peptide.

25. A storage medium having code for causing a processor to score peptide matches, the storage medium comprising:

code adapted to provide information associated with a first peptide and a second peptide;

code adapted to generate a stochastic model based on one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide;

code adapted to calculate a first probability that the first peptide matches the second peptide, based on the stochastic model;

code adapted to calculate a probability that the first peptide does not match the second peptide, based on the stochastic model; and

code adapted to score a match between the first peptide and the second peptide based at least in part on the ratio between the first probability and the second probability.

26. The storage medium according to claim 25, where

the first peptide is an experimental peptide; and

the second peptide is a candidate peptide, where the candidate peptide is selected from a group consisting of experimental peptides, theoretical peptides, and a library of peptides.

27. The storage medium according to claim 25, where the one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide comprise at least one of:

mass error;

charge state;

amino acid composition;

missed cleavage;

elution time;

protein/peptide modification;

mass spectrum peak intensity

mass spectrum signal to noise ratio;

mass spectrum signal quality indicator;

statistics derived from a database; and

any observable or derivable characteristics.

28. The storage medium according to claim 25 further comprising code adapted to determine the probability distributions for the one or more match characteristics.

29. The storage medium according to claim 28 further comprising code adapted to determine an empirical probability distribution for the one or more match characteristics

based on matches between experimental data for known peptides and peptides in a peptide database.

30. The storage medium according to claim 25 further comprising code adapted to adjust the stochastic model and a plurality of parameters associated with the stochastic model based on a learning data set, where the learning data set comprises a plurality of peptides that have been identified or a set of known protein standards.

31. The storage medium according to claim 25 further comprising code adapted to generate an output, where the output comprises at least one of:

a match score associated with the second peptide, where the match score comprises at least one of

a likelihood ratio, where the likelihood ratio is the ratio between the first probability and the second probability;

a log-likelihood, where the log-likelihood is the logarithm of the likelihood ratio;

the likelihood ratio divided by the length of the first peptide measured in amino acids;

the log-likelihood divided by the length of the first peptide measured in amino acids; and

the log-likelihood divided by the logarithm of the length of the first peptide measured in amino acids;

a Z-score associated with the match score;

a p-value associated with the match score;

biological information associated with the first peptide; and

biological information associated with the second peptide.

32. The storage medium according to claim 25, where a theoretical fragmentation spectrum is provided for the second peptide.

33. The storage medium according to claim 32, where the theoretical fragmentation spectrum includes masses corresponding to fragment isotopes.

34. The storage medium according to claim 25 further comprising code adapted to filter the second peptide based on at least one of:

the taxonomy of the protein that the second peptide belongs to;

the isoelectric point of the protein that the second peptide belongs to;

the molecular weight of the protein that the second peptide belongs to;

a non-symmetric mass window; and

a set of possible masses made of the union of a plurality of mass intervals.

35. The storage medium according to claim 25 further comprising

code adapted to provide a physical sample of the first peptide and biological information associated with the first peptide; and

code adapted to provide a physical sample of the second peptide and biological information associated with the second peptide.

36. A system for scoring a match between a first peptide and a second peptide, the system comprising:

means for generating a stochastic model based on one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide;

means for calculating a first probability that the first peptide matches the second peptide, based on the stochastic model;

means for calculating a probability that the first peptide does not match the second peptide, based on the stochastic model; and

means for scoring a match between the first peptide and the second peptide based at least in part on the ratio between the first probability and the second probability.

37. The system according to claim 36, where

the first peptide is an experimental peptide; and

the second peptide is a candidate peptide, where the candidate peptide is selected from a group consisting of experimental peptides, theoretical peptides, and a library of peptides.

38. The system according to claim 36, where the one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide comprise at least one of:

mass error;

charge state;

amino acid composition;

missed cleavage;

elution time;

protein/peptide modification;

mass spectrum peak intensity

mass spectrum signal to noise ratio;

mass spectrum signal quality indicator;

statistics derived from a database; and

any observable or derivable characteristics.

39. The system according to claim 36 further comprising means for determining the probability distributions for the one or more match characteristics.

40. The system according to claim 39 further comprising means for determining an empirical probability distribution for the one or more match characteristics based on matches between experimental data for known peptides and peptides in a peptide database.

41. The system according to claim 36 further comprising means for adjusting the stochastic model and a plurality of parameters associated with the stochastic model based on a learning data set, where the learning data set comprises a plurality of peptides that have been identified or a set of known protein standards.

42. The system according to claim 36 further comprising means for generating an output, where the output comprises at least one of:

a match score associated with the second peptide, where the match score comprises at least one of

a likelihood ratio, where the likelihood ratio is the ratio between the first probability and the second probability;

a log-likelihood, where the log-likelihood is the logarithm of the likelihood ratio;

the likelihood ratio divided by the length of the first peptide measured in amino acids;

the log-likelihood divided by the length of the first peptide measured in amino acids; and

the log-likelihood divided by the logarithm of the length of the first peptide measured in amino acids;

a Z-score associated with the match score;

a p-value associated with the match score;

biological information associated with the first peptide; and

biological information associated with the second peptide.

43. The system according to claim 36, where a theoretical fragmentation spectrum is provided for the second peptide.

44. The system according to claim 43, where the theoretical fragmentation spectrum includes masses corresponding to fragment isotopes.

45. The system according to claim 36 further comprising means for filtering the second peptide based on at least one of:

the taxonomy of the protein that the second peptide belongs to;

the isoelectric point of the protein that the second peptide belongs to;

the molecular weight of the protein that the second peptide belongs to;

a non-symmetric mass window; and

a set of possible masses made of the union of a plurality of mass intervals.

46. The system according to claim 36 further comprising

means for providing a physical sample of the first peptide and biological information associated with the first peptide; and

means for providing a physical sample of the second peptide and biological information associated with the second peptide.

47. A system for scoring a match between a first peptide and a second peptide, the system comprising:

a first calculation module that calculates a first probability that the first peptide matches the second peptide, based on the stochastic model;

a second calculation module that calculates a probability that the first peptide does not match the second peptide, based on the stochastic model; and

a scoring module that scores a match between the first peptide and the second peptide based at least in part on the ratio between the first probability and the second probability.

48. The system according to claim 47, where

the first peptide is an experimental peptide; and

the second peptide is a candidate peptide, where the candidate peptide is selected from a group consisting of experimental peptides, theoretical peptides, and a library of peptides.

49. The system according to claim 47, where the one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide comprise at least one of:

mass error;

charge state;

amino acid composition;

missed cleavage;

elution time;

protein/peptide modification;

mass spectrum peak intensity

mass spectrum signal to noise ratio;

mass spectrum signal quality indicator;

statistics derived from a database; and

any observable or derivable characteristics.

50. The system according to claim 47 further comprising a probability module that determines the probability distributions for the one or more match characteristics.

51. The system according to claim 50 further comprising an empirical module that determines an empirical probability distribution for the one or more match characteristics based on matches between experimental data for known peptides and peptides in a peptide database.

52. The system according to claim 47 further comprising an adjustment module that adjusts the stochastic model and a plurality of parameters associated with the stochastic

model based on a learning data set, where the learning data set comprises a plurality of peptides that have been identified or a set of known protein standards.

53. The system according to claim 47 further comprising an output module that generates an output, where the output comprises at least one of:

a match score associated with the second peptide, where the match score comprises at least one of

a likelihood ratio, where the likelihood ratio is the ratio between the first probability and the second probability;

a log-likelihood, where the log-likelihood is the logarithm of the likelihood ratio;

the likelihood ratio divided by the length of the first peptide measured in amino acids;

the log-likelihood divided by the length of the first peptide measured in amino acids; and

the log-likelihood divided by the logarithm of the length of the first peptide measured in amino acids;

a Z-score associated with the match score;

a p-value associated with the match score;

biological information associated with the first peptide; and

biological information associated with the second peptide.

54. The system according to claim 47, where a theoretical fragmentation spectrum is provided for the second peptide.

55. The system according to claim 54, where the theoretical fragmentation spectrum includes masses corresponding to fragment isotopes.

56. The system according to claim 47 further comprising a filter module that filters the second peptide based on at least one of:

the taxonomy of the protein that the second peptide belongs to;

the isoelectric point of the protein that the second peptide belongs to;

the molecular weight of the protein that the second peptide belongs to;

a non-symmetric mass window; and

a set of possible masses made of the union of a plurality of mass intervals.

57. The system according to claim 47 further comprising

a first provider module that provides a physical sample of the first peptide and biological information associated with the first peptide; and

a second provider module that provides a physical sample of the second peptide and biological information associated with the second peptide.

58. A peptide-matching method for diagnosing diseases, the method comprising:

providing a first peptide and a second peptide, where the first peptide is associated with at least one disease and the second peptide is not associated with the at least one disease;

generating a stochastic model based on one or more match characteristics associated with the first peptide, the second peptide and at least one fragment of the first peptide or the second peptide;

calculating a first probability that the first peptide matches the second peptide, based on the stochastic model;

calculating a probability that the first peptide does not match the second peptide, based on the stochastic model;

scoring a match between the first peptide and the second peptide based at least in part on the ratio between the first probability and the second probability; and

making diagnosis associated with the at least one disease based on the scored match between the first peptide and the second peptide.